MACHINE LEARNING AND CIVIL WAR: INVESTIGATING TREE-BASED MODELS FOR PREDICTING INTRASTATE VIOLENCE

by
Hilaire Meire

A capstone submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Government Analytics

Baltimore, Maryland
December 2017

Abstract

This study's aim is to improve the forecasting of civil war and examine the practical utility of using machine learning techniques in this effort. Specifically, this study investigates a variety of sampling methods used to construct useful models from imbalanced data, the algorithm used to construct these models, and which of the models built by previous scholars is the most useful for prediction when different sampling procedures algorithms are applied. This study finds that up-sampling and SMOTE sampling generally improve model performance, that tree-based ensemble methods generally perform significantly better than logistic regression and that of these ensemble methods Extreme Gradient Boosting generally performs the best, and that the previous model constructed by Collier & Hoeffler performs extremely well, especially when combined with sampling procedures and tree-based ensemble methods.

Table of Contents

1. Introduction

The causes of civil war have been the subject of a great deal of scholarly debate of the last 15 years. Primarily beginning with the work of Paul Collier and Anke Hoeffler[1] on economic motivations for intrastate violence and that of James Fearon and David Laitin[2] on the role of state capacity in civil war outbreak, the area of most discussion has been whether or not economic or political motivations are the true cause of civil war. However, this debate has not focused on attempting to use their findings to forecast future wars or to test how well their findings would predict past conflicts, instead focusing entirely on using linear statistical models to find the most statistically significant predictors. This lack of focus on predictive power and use of explicitly linear models in an attempt to model a non-linear relationship has led to a large body of literature that is of limited utility to policy makers given its methodological flaws and high specificity[3]. In addition, the capabilities of new machine learning models have advanced significantly since the debate on the causes of civil war began, but for the most part scholars have been slow to begin utilizing these new methods.

This study will attempt to address the weakness in the existing literature by utilizing predictive models in an attempt to forecast civil war, and testing

[1] Paul Collier, and Anke Hoeffler. "Greed and Grievance in Civil War." Oxford Economic Papers 56, no. 4 (2004): 536.; See also Paul Collier, and Anke Hoeffler. "On the Economic Causes of Civil War." Oxford Economic Papers 50, no. 4 (1998), 563.

[2] James D. Fearon, and David D. Laitin. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97, no. 1 (February 2003), 75.

[3] Michael D. Ward, Brian D. Greenhill, and Kristin M. Bakke. "The perils of policy by p-value: Predicting civil conflicts." Journal of Peace Research 47, no. 4 (2010), 363-375.

whether non-parametric tree-based models are better predictors of civil war than parametric methods such as logistic regression. It should be noted that there is a considerable difference between mathematical models used for explanations of a phenomenon and those used for prediction. The purpose of this study is not to craft a better explanation of civil war than Collier and Hoeffler or Fearon and Laitin, but to use the existing scholarship as a guide for crafting models that are better at predicting whether or not a civil war will occur. The use of machine learning to forecast whether or not a country may be about to descend into civil war is a powerful tool for policy makers and military leaders, and could be used as a means of concentrating efforts to prevent the predicted civil war from coming to fruition or to begin steps to mitigate the damage and end the war more quickly.

As stated above, machine learning is a growing field which is beginning to make inroads into the study of politics and political phenomena. A fairly digestible definition of machine learning is the use of an algorithm that can recognize patterns without being explicitly programmed. A simple practical example might look something like this: a bank wants to know whether or not someone will default on a loan, but have too many loan applications to be processed by a human, and believe a computer might pick up patterns that a human would miss. To solve this problem, the bank uses an algorithm that will classify whether or not an applicant will default on their loan based on a variety of inputs such as age, credit score, employment status, and education. The bank will first train the model to spot patterns on observations with a known outcome, in this case loans that have been issued and were either paid back or went into default. The bank will then use this trained algorithm to predict unseen data, and use its predictions as a guide for whether it should or should not issue the loan.

There is a small and developing body of existing scholarship on the application of machine learning to the study of civil war[4][5][6]. Previous works have echoed the criticisms of existing scholarship that have been covered here. However, these works have been primarily driven by a single question, "Is machine learning capable of predicting civil war?", and the findings have indicated a preliminary "yes"[7][8]. This study will go beyond that, and attempt to find the general type of algorithm which predicts civil wars best. The number of machine learning algorithms is large, and finding the ones which are capable of predicting such a complex phenomenon is an important task if this methodology is to be applied to the study of civil war.

2. *Literature Review*

The goal of this paper is to test different machine learning models for use in forecasting civil war outbreak. As a result, this literature review will cover two general bodies of political science literature; the body of work dealing with the causes of civil war and the work dealing with forecasting civil war with parametric and non-parametric machine learning techniques. The latter is still

---

[4] Jack Goldstone, Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward. "A global model for forecasting political instability." American Journal of Political Science 54, no. 1 (2010), 190.
[5] Chris Perry. "Machine Learning and Conflict Prevention: A Use Case." Stability: International Journal of Security and Development 2, no. 3 (2013), 1.
[6] David Muchlinski, David Siroky, Jingrui He, and Matthew Kocher. "Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data." Political Analysis 24, no. 1 (2015): 87.
[7] Perry, "Machine Learning", 17-18.
[8] Muchlinski, "Comparing Random Forests", 87.

very much a developing body of work, and tends to sit somewhere between the disciplines of political methodology and computer science.

The study of civil war outbreak has, for at least the least 13-15 years, been tightly focused around the fundamental question of whether political grievance against the controlling institution or the economic needs and desires of the rebel groups. This debate is typically labelled the "greed versus grievance" debate, in reference to the seminal article by Collier and Hoeffler, "Greed and Grievance in Civil War"[9]. This article, published in 2004, maintains that rebel greed, rather than political grievance, appears to be the primary cause of civil war. Examining civil wars from 1960-1999, the authors construct a series of logistic regression models to test whether greed or grievance based variables are better predictors. In their best fitting model, they find that the most significant predictor of whether a country will experience a civil war is primary commodity exports as a percentage of GDP. The authors conclude that this is a result of these exports providing a relatively easy means of financing the rebellion. Overall, the model they construct using only economic variables leads to a far better fit than the model constructed using only economic variables. It should be noted that the authors admit that political grievance variables, such as ethnic tension and political repression, are difficult to quantify.

[9] Collier and Hoeffler, "Greed and Grievance", 563; See also Collier and Hoeffler, "On the Economic Causes", 536. which covers similar topics and began a more in-depth discussion of economic motivations in civil war, although it is less influential.

The findings of Goldstone et al in their 2010 paper run counter to the assertions of Collier and Hoeffler[10]. Goldstone et al find that four variables: regime type, infant mortality rate, the presence of armed conflict in 4 or more bordering countries, and state-led discrimination, are able to accurately predict whether or not a civil war will break out in over 80% of the cases tested, with regime type being the most powerful predictor. Their methodology is more precisely tuned to forecasting rare events in cases where confounding variables are highly likely to be present, and the utilize a more practical, machine learning oriented test for evaluating their models. Rather than evaluating these models based mainly on R^2 values and measures of statistical significance (although the latter is present and considered), the authors test their models on randomly selected sample data not used in its construction. However, the authors do note that evaluating models based on predictive power is fundamentally different than evaluating them on fit, and thus not comparable in a strict sense. In the end, they conclude that regime type is the most powerful predictor of civil war onset (in addition to other forms of political instability), and that other political and social variables become almost insignificant when regime type is taken into account.

Buhaug et al are also critical of Collier and Hoeffler, and assert that prior measures of economic and ethnic fractionalization, such as the Gini coefficient and Ethno-Linguistic Fractionalization metric, do not effectively capture the underlying causes of economic and ethnic exclusion and thus may create

[10] Goldstone et al, "A Global Model", 190.

unreliable models[11]. The authors posit that economic inequality as measured by the Gini coefficient is useful primarily for evaluating economic inequality between individuals (termed vertical inequality), rather than inequality between groups as a whole (horizontal inequality). As civil wars are waged by groups acting collectively rather than individuals, horizontal inequality is seen as a more appropriate indicator for studying their outbreak than vertical inequality. They also maintain that there are several types of intrastate conflict: ethnically based territorial, ethnically based governmental conflict, and non-ethnic conflict. The authors find that separating the three conflict types, utilizing their higher-level group based indicators of economic inequality, and focusing primarily on the size of the largest oppressed group as a fraction of the total country population, it is clear ethnic groups which are far poorer than the national average for a group are more likely to trigger ethnic based territorial conflict, and that the presence of a large, oppressed ethnic group is more likely to lead to ethnically based governmental conflict.

Fearon and Laitin fall somewhat outside the greed versus grievance debate. Their 2003 paper, "Ethnicity, Insurgency, and Civil War"[12], has proved to be almost as influential and Collier and Hoeffler's, and maintains certain criteria, such as a poverty, rugged terrain, and a weak central government, are conditions highly favorable to rebellion, and their presence increases the likelihood of civil war outbreak. They find that these variables are also far better predictors of civil

[11] Halvard Buhaug, Lars-Erik Cederman, and Kristian Skrede Gleditsch. "Square pegs in round holes: Inequalities, grievances, and civil war." International Studies Quarterly 58, no. 2 (2014), 418.

[12] Fearon and Laitin, "Ethnicity", 75.

war outbreak than state discrimination, ethnic and religious tension, and economic inequality. The authors note that conditions that make it easier for rebels to hide from government forces (the conditions that favor civil war outbreak) make it possible for a small group of committed rebels to wage a sustained campaign against a larger government, and make it more difficult to stop a potential rebel group.

Fearon and Laitin's findings are supported by Reagan and Norton[13]. In "Greed, Grievance, and Mobilization in Civil Wars", they argue that political grievance is what initiates civil wars, but that economic motivations become more important for the movement to remain viable. Similar to what Fearon and Laitin propose, they also find that a government's ability to react more effectively to lower level forms of political instability decreases the likelihood of civil war breaking out. Finally, they find that in order to be effective means of funding a rebellion, natural resources must be easily lootable, transportable, and saleable for a rebel group. This means resources that require a great deal of infrastructure to extract or are very large and cumbersome, such as timber and oil, are not especially useful to rebel groups as a source of funding, while resources such as precious stones and especially drugs are much more viable.

Hegre and Sambanis find a variety of significant relationships in their 2006 study "Sensitivity Analysis of Empirical Results on Civil War Onset", many

[13] Patrick M. Regan, and Daniel Norton. "Greed, grievance, and mobilization in civil wars." Journal of Conflict Resolution 49, no. 3 (2005), 319.

of which support a generally economic impetus for civil war[14]. They find that countries with a large population, low per capita GDP, and sluggish annual GDP growth are more likely to experience civil war, which is consistent with the bulk of the civil war literature. However, they also find support for Fearon and Laitin's assertion that rough terrain and state capacity (here represented as military size) increase the likelihood of civil war. In addition, they find that being located in a region with many other states that have experienced recent warfare and are generally undemocratic also increases the likelihood.

Ward et al set the stage for later discussion of the use of machine learning methodologies in the study of civil wars with their critique of what they saw as the over-reliance on descriptive statistics and statistical significance in the field[15]. They maintain that previous researchers had focused too much on analyzing the statistical significance of the variables they had available, rather than finding or engineering more useful variables, similar to what Buhaug et al did. They argue that both Fearon and Laitin and Collier and Hoeffler's models lack predictive power, and that while they were created to be descriptive, not predictive, this fact alone does not make them better models or better indicators of the true causes of civil war (it should be noted that when tested for their predictive power, Collier and Hoeffler's model is a better predictor). The authors close by warning that basing policy on models that have been constructed primarily to be explanatory and to find statistically significant variables is likely to have undesirable

[14] Havard Hegre, and Nicholas Sambanis. "Sensitivity Analysis of Empirical Results on Civil War Onset." Journal of Conflict Resolution 50, no. 4 (August 2006), 508.
[15] Ward, "Perils", 363.

outcomes and should be avoided, and that further research should be more aware of the predictive power of the models and place less emphasis on finding the highest p-value.

As stated above, work on using machine learning methods to forecast civil war is a developing field. All of the papers previously described, with the exception of Goldstone et al, have used fairly standard social science statistics as their primary methodologies, primarily logistic regression in this area. This is certainly a useful explanatory tool when examining the underlying causes of political phenomena, it does not fit the need for a reliable way of detecting potential conflicts early and reacting to them. The subsequent works deal with this topic in more detail.

In “Machine Learning and Conflict Prevention: A Use Case”, Perry examines the practical utility of machine learning methods for predicting the outbreak of conflict[16]. Using data covering Africa at the subnational level (provinces and cities), the variables used to construct the models are GDP, population, the presence of diamonds and petroleum deposits, and the percentage of children who are underweight. Several models were conducted using the naïve Bayes and random forest algorithms. The first two models used only previous incidence of battles as input variables, while the second two used the full dataset. Finally, a random forest model was conducted to predict the number of battles in a given area. The classification models exhibit high overall accuracy, which is

[16] Perry, “Machine Learning”, 1-17.

unsurprising. The actual occurrence of a battle is rare, and thus it seems intuitive that the model would correctly predict the majority of cases correctly simply by predicting false in virtually all cases. The author concludes by noting that this new capability represents a step forward for policy makers, and that while his models may not seem to be groundbreaking, the capability the use is. He outlines five steps forward, which are as follows: examine whether a more localized or more global scope is better for the task, look into newer tools for building larger versions of these models at faster speeds, such as cloud computing and parallel processing, examine the effect of time on the predictive power of the models in a more systematic and robust way, locate better data for the construction of future models, and create a more detailed outcome categorization that goes beyond a simple binary indicator of whether or not a battle happened or a straight count of the number of battles which occurred.

Muchlinski et al compare the predictive power of several different forms of logistic regression (L1-regularized, Rare Events, and traditional logistic regression) to that of random forest models[17]. The authors here advocate using cross-validation in the construction and evaluation of a model. This consists of shuffling the dataset and iteratively reconstructing the model using different folds of the data, so that each observation is used both as data used to construct the model and used to test the model at different times[18]. The overall evaluation

---

[17] Muchlinski, "Comparing Random Forest", 87-103.

[18] Colaresi, Michael, and Zuhaib Mahmood. "Do the robot: Lessons from machine learning to improve conflict forecasting." Journal of Peace Research 54, no. 2 (2017), 193-214. also advocate for cross-validation and an iterative framework for constructing predictive models. While their framework is somewhat intuitive (use domain expertise to select model parameters, construct the model, test it against existing data, evaluate flaws, then use domain expertise to

metric of the model is an aggregation of the different times the model was run. Utilizing this technique for the three types of logistic regression and the random forest model, the researchers find that the random forest model is far better at predicting whether or not a civil war will break out. It is also worth noting that the least powerful predictor variable in the random forest model is primary commodity exports, while the most powerful are GDP growth and GDP per capita. The authors conclude that random forests appear to be a much more accurate predictor than logistic regression, but note that logistic regression may prove superior in cases where the relationship between independent and dependent variables is more linear.

Reviewing the existing literature reveals that there is an apparent hole in the research currently being conducted into civil wars. Previous research has relied primarily on using descriptive statistics to find statistically significant variables in explanatory models, and the existing research on predictive models is rather sparse. In addition, the research into predictive models seems to have been primarily utilized random forests, which are only one algorithm in the machine learning toolbox. This study will attempt to address this gap by utilizing different algorithms, and evaluating their power against both random forests and logistic regression.

---

attempt to address these flaws and go through the process again), it runs counter to the more p-value driven process typically used in the social sciences. It has not been elaborated on it in more detail because it is so intuitive, but it is worth noting that developing a standard workflow and advocating for the use of cross-validation is still seen as being necessary in this sub-field as recently as 2017.

3. Data and Methods

The goal of this study is to test how well previous research fares in predicting civil war outbreak, particularly when tree-based non-parametric models were used. This study tests six predictive algorithms: logistic regression, random forests, boosted classification trees, bagged classification trees, stochastic gradient boosting, and extreme gradient boosting. Of these, logistic regression is the only parametric model. Data utilized are the same data used by Hegre & Sambanis[19] and Muchlinski et. al[20]. The data is recorded for every formally recognized country in the world for each year, 1945-2000. The outcome variable (the variable to be predicted) is whether or not a civil war incident occurred in that year. An incident of civil war in this dataset is defined as a conflict between at least two intrastate actors that results in at least 1000 battle deaths within the year. Models in this dataset were constructed in R using the Caret machine learning library. K-fold cross validation with 10 folds was utilized.

Civil war is a very rare event, and as a result these data are highly unbalance (roughly 1% of the observations in the dataset experienced civil wars). In order to correct for this, four different sampling methods were tests: down-sampling and up-sampling, which are bootstrap methods which under-sample the most common class and over-sample the uncommon class, and the ensemble

[19] Hegre and Sambanis, "Sensitivity Analysis", 508.
[20] Muchlinski, "Comparing Random Forest", 4-5.

sampling methods of Synthetic Minority Over-Sampling Technique (SMOTE)[21] and Random Over-Sampling Examples (ROSE)[22], both of which generate additional positive class examples algorithmically. Each model was also run with no sampling for a baseline.

In machine learning, there are several different metrics which can be used to assess the performance. Metric selection is a highly subjective process, and is extremely dependent on the purpose of the model. This study relies upon three metrics, those of the Area Under the Curve (AUC), Sensitivity, and Specificity. The area under the curve plots the true positive rate (in this case incidents of war correctly classified as such by the model) against the false positive rate (incidents of war classified as incidents of peace), and calculates the area underneath ranging from 0-1. Evaluating the AUC score is extremely similar to traditional academic grades, with .9 – 1 being a very powerful model, .8-.89 being a good model, 0.7-0.79 being a fair model, 0.6 – 0.69 being a weak model, and anything lower being useless model. Sensitivity is the percentage of the positive class (here an incident of civil war) correctly predicted and specificity is the percentage of the negative class (here peace) correctly predicted.

It should be mentioned at this point that this study is an extension of the work done by Muchlinski et al. While their study only compared logistic regression to

[21] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002), 321-357.
[22] Menardi, Giovanna, and Nicola Torelli. "Training and assessing classification rules with imbalanced data." Data Mining and Knowledge Discovery (2014), 10.

random forests[23], this study attempts to go further at test a variety of tree-based algorithms. Their study also utilized down-sampling only the case of their final model, not for every model constructed, and does not test various sampling methods. Finally, Muchlinski et al construct their models using peace as the positive class (the class the model is trying to predict)[24].

In order to gain a general understand of which type of variables were the best predictors in isolation, models were constructed using only variables in each of the following categories: economic, governmental, demographic, and developmental. Economic variables included GDP per capita, GDP growth, primary commodity exports as a percentage of GDP, etc. Governmental variables included regime type, military manpower, and regime durability among others. Demographic variables included ethnic and religious fractionalization, while developmental variables included things such as infant mortality rate and literacy rate. The specific variables for each model are listed below

Fearon & Laitin[25]: History of Conflict, Log Population, Log GDP, Rough Terrain, Noncontiguous, Oil Exports (% GDP), New State, Political Instability, Polity IV Score, Ethnic Fractionalization, Religious Fractionalization

Collier & Hoeffler[26]: Primary Commodity Exports (% GDP) Squared, Log GDP, Annual Change in GDP, History of Conflict, Rough Terrain, Ethnic Fractionalization, Population Density, Log

[23] Muchlinski, "Comparing Random Forest", 1.
[24] This is a rather subtle error to spot, and is a bug in the authors' code resulting from a somewhat strange way the Caret library for R handles classes. Without specification, it takes the level with either the lowest number of the first letter as the positive class.
[25] Fearon and Laitin, "Ethnicity", 84.
[26] Collier and Hoeffler, "Greed and Grievance", 573.

Population, Cold War Era, Secondary Education Rate

Hegre & Sambanis[27]: Log Population, Log GDP, Political Instability, Regulation of Political Participation, Middle East or North Africa, Durability Metric, GDP Growth, Anocracy, Partial Freedoms, Neighbor at War, Rough Terrain, Polity IV Score Squared, New State, Median Regional Polity Score (Polity II), Ethnic Dominance Metric, Military Size (Personnel), Western Europe or US, Number of Neighbors at War, Presidential System

Economic: Agriculture Exports (% GDP), Exports (% GDP), Oil and Fuel Exports (% of Material Exports), Annual GDP Growth, Manufactured Exports (% of Material Exports), Average Neighbor Log GDP per Capita, Oil (% of GDP), Primary Commodity Exports (% GDP) Squared, Trade (% GDP)

Governmental: Regime Age, Anocracy, Size of Government Army (1985), Autocracy Score, Annual Change, Autocracy Index (Polity IV), de Facto Autonomous Regions, Centralized State (Polity III), Democracy Index (Polity IV), Democracy Index Change (Polity 98), Regime Durability, Federal State (Polity III), Regime Type, Incumbent Advantage, Majoritarian System, Military Size (Personnel), Polity Type, Neighbors' Median Polity Index (Polity II), New State, Annual Polity Change, Competitiveness of Political Participation for Non-Elites, Regulation of Political Participation, Log of Voting Population to Opposition's Share of Votes Cast, Partially Free Polity, Polity IV Index, Polity Change (Polity 98), Political Competition (Polity IV), Presidential System, Median Regional Polity (Polity II), Semi-Federal State (Polity III)

Demographic: Linguistic Heterogeneity, Racial Heterogeneity, Religious Heterogeneity, Ethnic Fractionalization, Ethnic Heterogeneity, Ethnolinguistic Diversity, Ethnic Dominance Metric, Number of Linguistic Groups, Population Share of Largest Ethnic Group,

[27] Hegre and Sambanis, "Sensitivity Analysis", 528.

Religious Fractionalization, Percent of Population in Second Largest Ethnic Group

Development: Illiteracy Rate, Infant Mortality Rate, Life Expectancy, Primary School Enrollment, Secondary School Enrollment

4. *Results*

Algorithm	Variable Type	Med AUC	Med Sensitivity	Med Specificity
	Economic	0.8157	0.7841	0.7375
	Governmental	0.6846	0.4553	0.6279
	Demographic	0.6144	0.4477	0.6226
	Developmental	0.6748	0.4492	0.6846
	Economic	0.9196	0.6909	0.9093
	Governmental	0.8212	0.2636	0.8982
	Demographic	0.6978	0.4076	0.8643
	Developmental	0.8449	0.6227	0.875
	Economic	0.9188	0.2492	0.1492
	Governmental	0.8423	0.2053	0.2544
	Demographic	0.7525	0.3977	0.3203
	Developmental	0.8792	0.275	0.2081
	Economic	0.8617	0.653	0.8955
	Governmental	0.7363	0.2712	0.8556
	Demographic	0.6593	0.422	0.8515
	Developmental	0.797	0.6121	0.8451
	Economic	0.9011	0.8008	0.8068
	Governmental	0.8474	0.5159	0.8483
	Demographic	0.7655	0.4424	0.7608
	Developmental	0.8663	0.7068	0.8215
	Economic	0.9219	0.6636	0.9006
	Governmental	0.8451	0.5333	0.8776

Demographic	0.7735	0.5068	0.8358
Developmental	0.8781	0.7341	0.796

Figure 1

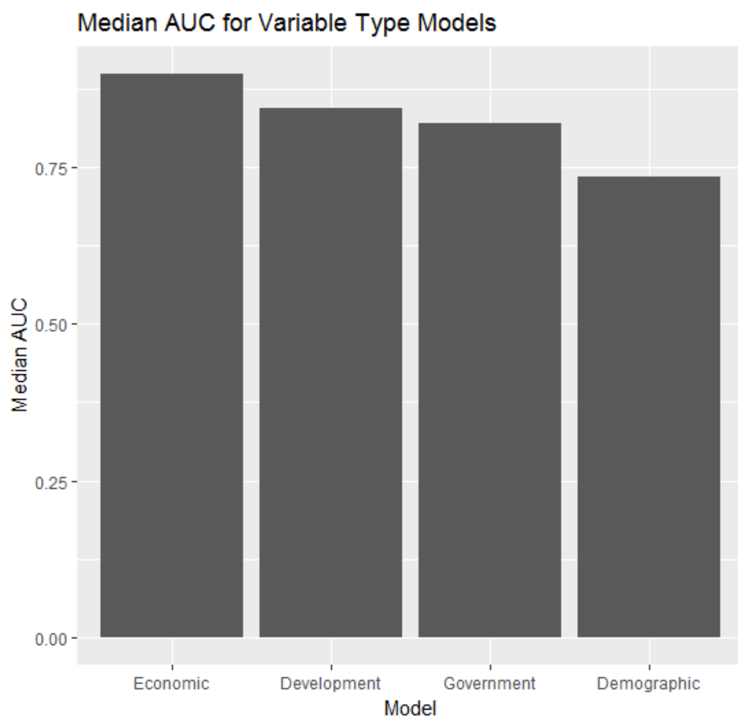
As can be seen here, economic variables are the strongest predictor, followed by governmental and developmental. Demographic variables are the weakest predictor. This generally lends support to Collier & Hoeffler's assertion that economic issues drive civil war[28]. However, it is important to note that prediction is fundamentally different from explanation, and it is not wise to draw conclusions about underlying cause from predictive models. It should also be noted that highest median specificity, meaning the percentage of wars correctly predicted, belongs to logistic regressions constructed with economic variables. However, an important caveat is that in some cases, models may predict a single class for every observation, and thus it is important to take all the metrics into account. A model that predicts a civil war will break out in every country every year will predict all civil wars, but clearly as useless as a model that predicts civil wars will never break out anywhere. This table also shows the power of tree-

[28] Collier and Hoeffler, "Greed and Grievance", 595.

based models compared to logistic regression, which performs worse for every model.

Figure 2

It is also important to note that Demographic variables are the worst category of predictor variables on their own. This shows that ethnic, religious, and linguistic fractionalization alone are not enough to accurately predict civil war outbreak. Similarly, while performing a good deal better than Demographic variables, Governmental variables perform significantly worse than Economic



variables, and are also not a powerful predictor of civil war outbreak on their own. Finally, the second most powerful category, Developmental, may be intuitively understood to be related to economic development (typically a wealth nation will not have a high infant mortality rate or a large portion of its population that cannot read).

The second step in the process was to test the variables utilized by previous scholarship to see how well they performed with different algorithms. The previous models tested were those constructed by Fearon & Laitin, Collier & Hoeffler, and Hegre

& Sambanis.

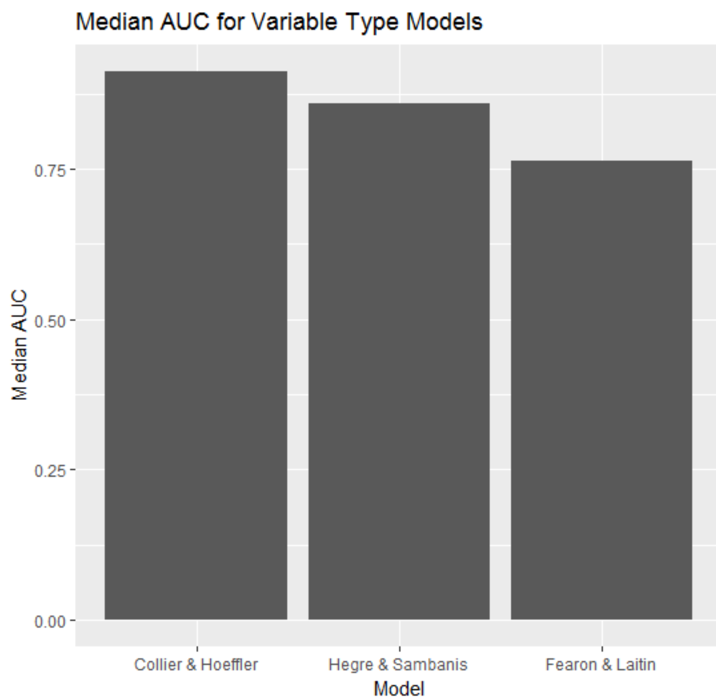
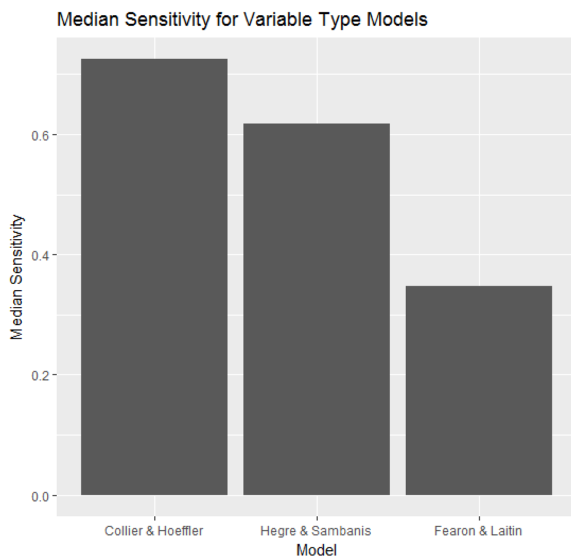


Figure 3

Algorithm	Previous Model	Median AUC	Median Sensitivity	Median Specificity
Logistic Regression	Fearon & Laitin	0.7634	0.5856	0.726
	Collier & Hoeffler	0.822	0.6985	0.7575
	Hegre & Sambanis	0.7947	0.6394	0.7327
Random Forest	Fearon & Laitin	0.7479	0.1432	0.9524
	Collier & Hoeffler	0.9133	0.725	0.9008
	Hegre & Sambanis	0.8666	0.4568	0.9095
Boosted Trees	Fearon & Laitin	0.7697	0.309	0.2972
	Collier & Hoeffler	0.9276	0.2864	0.1511
	Hegre & Sambanis	0.8784	0.4167	0.1439
Bagged CART	Fearon & Laitin	0.7023	0.0848	0.9741
	Collier & Hoeffler	0.8486	0.7182	0.8971
	Hegre & Sambanis	0.7674	0.5689	0.874
Stochastic Gradient Boosting	Fearon & Laitin	0.7789	0.5523	0.793
	Collier & Hoeffler	0.922	0.7742	0.8914
	Hegre & Sambanis	0.8659	0.7235	0.8354
Extreme Gradient Boosting	Fearon & Laitin	0.7766	0.4735	0.8256
	Collier & Hoeffler	0.9297	0.8106	0.8824
	Hegre & Sambanis	0.8735	0.65565	0.8321

Figure 4



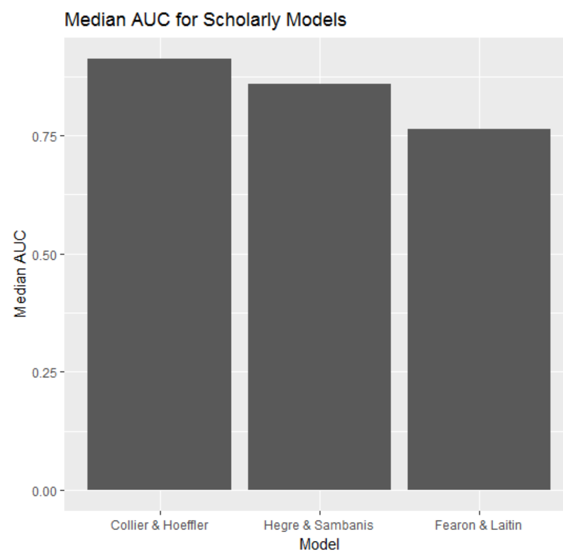


Figure 5

Here, we can see that Collier & Hoeffler's model has a great deal of predictive power, while Fearon and Laitin's has the least. Overall, variables selected by Collier & Hoeffler used with extreme gradient boosted trees have the highest median AUC at 0.9207 and the highest sensitivity with 81.06% of war incidents predicted correctly. The specificity (the number of peace years correctly predicted to be peaceful) is also very high, at 88.24% correct.

Examining the results shows that a few algorithms truly lag behind in terms of predictive power. Overall the algorithm that made the strongest predictions was Extreme Gradient Boosting, and the least powerful was Logistic Regression. Extreme Gradient Boosting, Boosted Classification Trees, and Stochastic Gradient Boosting all performed comparably, with Random Forests lagging slightly behind and Bagged Trees following Random Forests. Clearly, boosting is a powerful technique to use in the prediction of civil war.

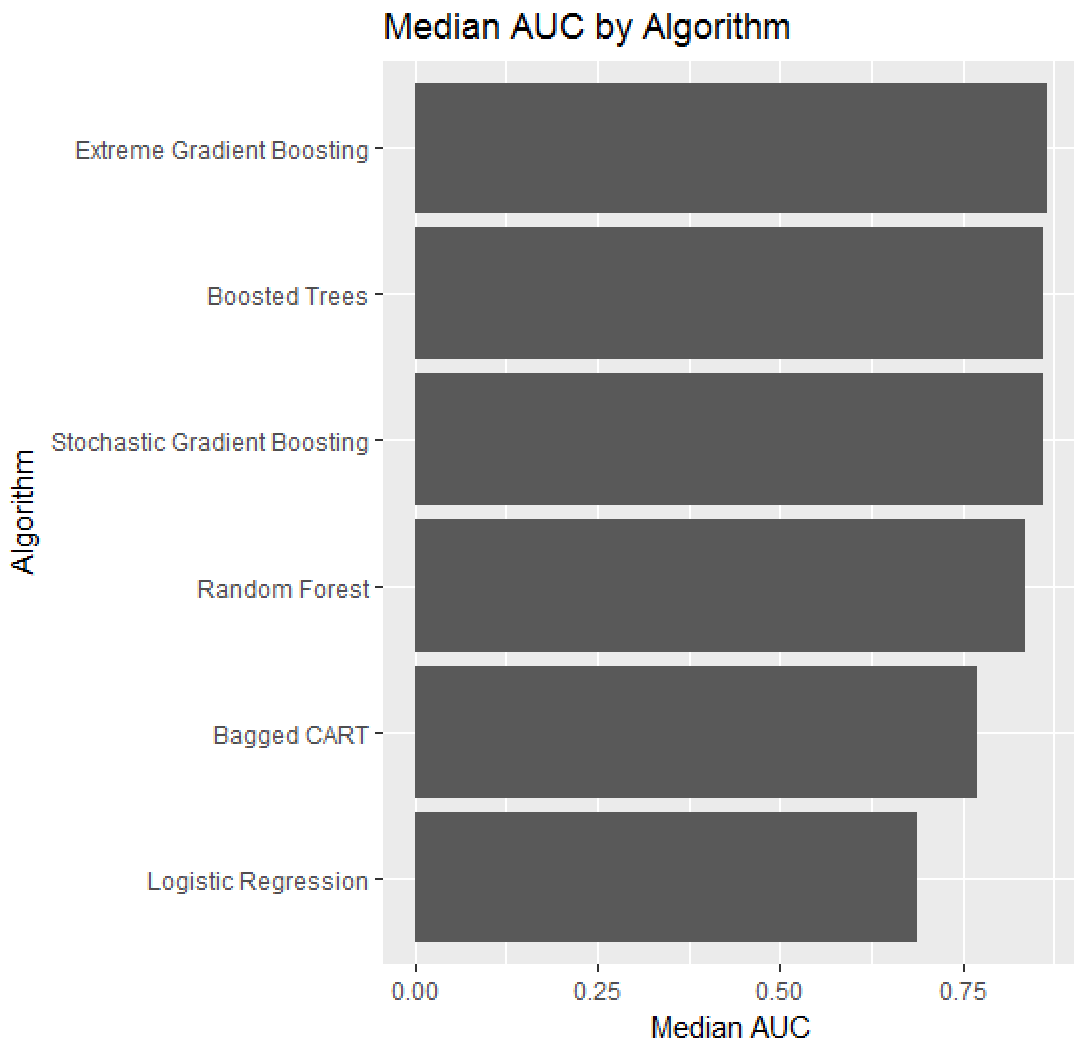
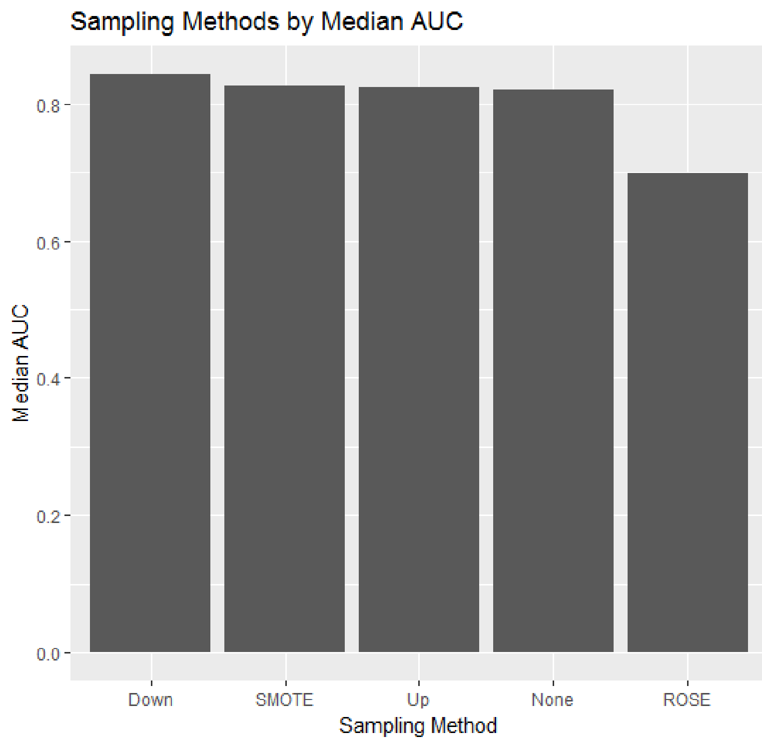


Figure 6

Sampling methods were also an area investigated by this study, as stated above. To reiterate, the sampling methods used were up-sampling (sampling examples of the positive class multiple times to increase its effect on the model), down-sampling (leaving out samples of the negative class to decrease its effect relative to positive examples), SMOTE and ROSE (ensemble methods which



algorithmically generate new examples of

Figure 7

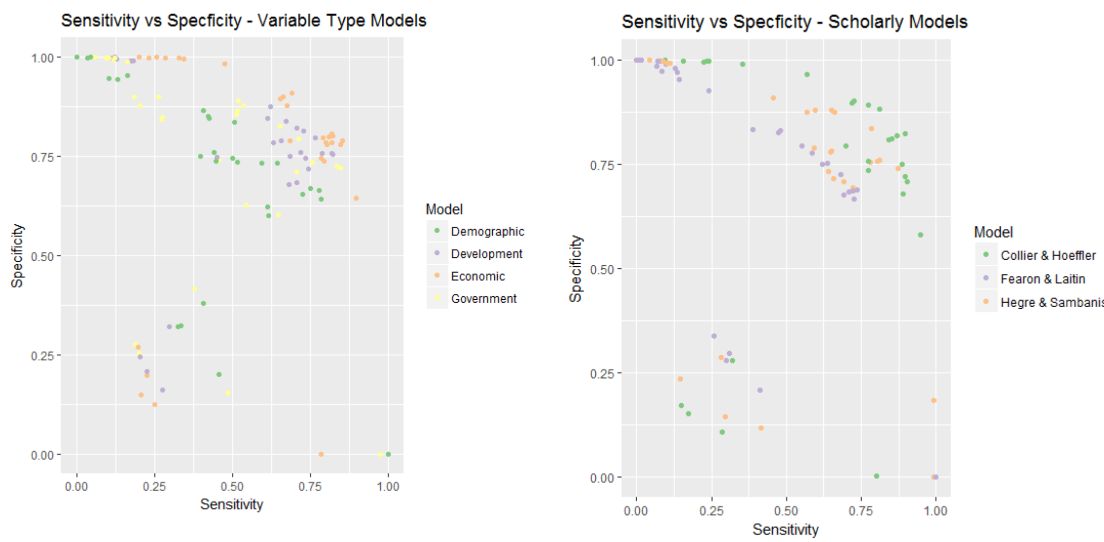
the positive class based on the mathematical properties of existing samples). Overall, all sampling methods performed relatively consistently, with ROSE being the only metric that did worse that the control group that used no sampling. Down-sampling performed the best overall, but does not appear in the top ten models by AUC, and only marginally outperformed SMOTE, which does appear in the top ten.

Type	Model	Sampling	AUC	Sensitivity	Specificity
Extreme Gradient Boosting	Collier & Hoeffler	None	0.9356	0.2409	0.9976
Boosted Trees	Collier & Hoeffler	Up	0.9337	0.2864	0.1081
Stochastic Gradient Boosting	Collier & Hoeffler	SMOTE	0.932	0.7742	0.8914
Random Forest	Economic	None	0.9314	0.1992	0.999
Extreme Gradient Boosting	Collier & Hoeffler	SMOTE	0.9309	0.8106	0.8824
Extreme Gradient Boosting	Collier & Hoeffler	Up	0.9297	0.5689	0.9647
Boosted Trees	Collier & Hoeffler	None	0.9281	0.7992	0.0021
Random Forest	Collier & Hoeffler	SMOTE	0.9279	0.725	0.9008
Boosted Trees	Collier & Hoeffler	SMOTE	0.9276	0.1727	0.1511
Extreme Gradient Boosting	Economic	None	0.9271	0.2583	0.9989

Figure 8

The ten models with the highest AUC are shown above. While the Extreme Gradient Boosted Collier and Hoeffler model with no sampling has the highest AUC at .9356, it is important to note that the specificity of the model is low, meaning it is likely to under-predict civil war outbreak. While it is possible to argue otherwise, it seems that the Extreme Gradient Boosted Collier and Hoeffler model that utilized SMOTE sampling is the most powerful model all around, as it boasts an AUC of over .9, a sensitivity of 0.8106 and a specificity of 0.8824. It should also be noted here that eight out of ten of the models that appear on this are Collier and Hoeffler, and the other two use only economic variables. Additionally, the algorithm that appears most often is Extreme Gradient Boosting, with random forests (the method utilized by Muchlinski et al) only appearing twice, and the first appearance under-predicting incidents of civil war.

As can be seen in the scatter plots below, when comparing models of each category, Collier & Hoeffler and Economic are the front-runners in this area as well. Points which appear closest to the upper-right corner have the highest combination of sensitivity and specificity. As was stated above, developmental variables come second for the variable type models. While they do perform slightly less well than economic variables, it is important to note that there are



fewer of them.

Figure 9

5. *Conclusion*

The results of this study indicate three things: non-parametric tree-based models perform significantly better in predicting civil war outbreak than logistic regression and more nuanced ensemble methods perform the best of those tree-based models, economic variables generally and Collier & Hoeffler's economic based model specifically are generally better predictors of civil war outbreak, and ROSE sampling is highly ineffective for prediction of this topic, while all other sampling methods performed comparably.

It is perhaps unsurprising that civil war is predicted better by models which are not strictly linear. However, this study has demonstrated just how large the gap is in this case, and also has significant implications for future research. While this study does not claim to provide any explanation as to the causes of civil war, it does suggest that expanding statistical techniques beyond parametric techniques can yield strong results. It seems plausible that future studies with the goal of explanation may do well to utilize tree regression rather than linear regression to yield more robust results. Similarly, it is simply not enough to rely on the Random Forest algorithm to outperform logistic regression, as it was routinely outperformed by Extreme Gradient Boosting and Stochastic Gradient Boosting.

As to the second conclusion of this study, it appears that, whatever the underlying cause of civil war, economic condition is the most powerful predictor of whether one will erupt in a given country during a given year. Using only economic variables, it is possible to construct a model with an AUC of greater than .9, which, as stated above, is quite a powerful model. Additionally, the previous work by Collier & Hoeffler turns out to have a great deal of predictive power, dominating the top ten most powerful models by AUC and in some cases having greater than .8 sensitivity and specificity simultaneously. While it may be possible to push the performance ceiling even higher using feature engineering or dimensionality reduction, this model performs extraordinarily well as it stands. This also demonstrates that previous, explanation-driven scholarship is useful in more prediction oriented research.

Regarding the issue of sampling, it seems surprising that models which use no sampling at all should outperform any sampling method, given the extreme imbalance in this dataset. This one example is certainly not enough to disregard to the ROSE sampling method entirely. However, future research may be needed to investigate its practical utility in the world of predictive modeling in political science.

As has been hinted, this study contains a number of interesting avenues for future research. The first and largest is the usefulness of artificial neural networks for predicting civil war. While these methods are computationally intensive, it is quite possibly the case they perform strongly enough that the trade-off is worth the increase in complexity and computational intensity. In a similar vein, dimensionality reduction and feature engineering should be investigated as a possible means of improving the ability to predict intrastate conflict. Finally, a thorough test of ROSE sampling across a number of topics in political science and specifically comparative politics could help in determining whether the method has any utility in this discipline. Given how frequently imbalanced data present themselves in political research, it will always be useful to have more ways of dealing with this imbalance. However, the poor performance of this sampling method in this study would seem to indicate there are areas in which it simply is not suited.

Ideally the techniques investigated in this study will help to create more robust early warning systems for civil conflict and be helpful in crafting policy to increase stability and reduce violence throughout the world. While the use of machine learning for public policy is still in its infancy, it shows a great deal of promise in changing the way policy is created.

References

1. Buhaug, Halvard, Lars-Erik Cederman, and Kristian Skrede Gleditsch. "Square pegs in round holes: Inequalities, grievances, and civil war." International Studies Quarterly 58, no. 2 (2014): 418-431.
2. Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.
3. Colaresi, Michael, and Zuhaib Mahmood. "Do the robot: Lessons from machine learning to improve conflict forecasting." Journal of Peace Research 54, no. 2 (2017): 193-214.
4. Collier, Paul, and Anke Hoeffler. "Greed and Grievance in Civil War." Oxford Economic Papers 56, no. 4 (2004): 536-95.
5. Collier, Paul, and Anke Hoeffler. "On the Economic Causes of Civil War." Oxford Economic Papers 50, no. 4 (1998): 563-73.
6. Fearon, James D. "Primary commodity exports and civil war." Journal of conflict Resolution 49, no. 4 (2005): 483-507.
7. Fearon, James D., and David D. Laitin. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97, no. 1 (February 2003): 75-90.
8. Goldstone, Jack A., Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward. "A global model for forecasting political instability." American Journal of Political Science 54, no. 1 (2010): 190-208.

9. Hegre, Havard, and Nicholas Sambanis. "Sensitivity Analysis of Empirical Results on Civil War Onset." Journal of Conflict Resolution 50, no. 4 (August 2006): 508-35.
10. Menardi, Giovanna, and Nicola Torelli. "Training and assessing classification rules with imbalanced data." Data Mining and Knowledge Discovery (2014): 1-31.
11. Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher. "Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data." Political Analysis 24, no. 1 (2015): 87-103.
12. Perry, Chris. "Machine Learning and Conflict Prevention: A Use Case." Stability: International Journal of Security and Development 2, no. 3 (2013): 1-18.
13. Regan, Patrick M., and Daniel Norton. "Greed, grievance, and mobilization in civil wars." Journal of Conflict Resolution 49, no. 3 (2005): 319-336.
14. Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. "The perils of policy by p-value: Predicting civil conflicts." Journal of Peace Research 47, no. 4 (2010): 363-375.

Curriculum Vita

Hilaire Meire graduated from Virginia Commonwealth University with a BA in Political Science (International Relations) and History, Magna Cum Laude, in 2014. He has been a student in the MS in Government Analytics at Johns Hopkins University since 2015. He is from Richmond, Virginia and was born in May, 1991.